

A Causal Theory of Modal Knowledge (Including Logical and Mathematical Knowledge)

Robert C. Koons
Department of Philosophy
University of Texas at Austin
koons@phil.utexas.edu
(512) 471-5530

August 9, 1999

1 Introduction

The notion of causality is absolutely central to recent philosophical work in semantics, the philosophy of mind and intentionality, epistemology, and philosophy of science. Work by Donnellan, Kripke [19], and Putnam [26] helped to make causal connections an indispensable part of our accounts of reference and signification. This in turn has generated causal theories of information and content ([9] and [13]). Causality has been pressed into service through the construction of aetiological accounts of teleology and proper function.[29][33][17] The Gettier problem led to the renaissance of causal and teleo-causal theories of knowledge by Goldman [14], Armstrong [1] Pollock [25], and Plantinga [24]. Causality is put to much work in recent theories of personal identity and of the nature of mental states (as in the functionalism of Lewis [21] and Putnam [26]). Causation continues to figure prominently in philosophy of science (e.g., Wesley Salmon's causal theory of evidence [27]).

However, the project of building a unified theory of intentionality and knowledge in causal (or teleo-causal) terms faces a major obstacle: accounting for our knowledge of *modal* facts, i.e., facts about necessity and possibility (including logical and mathematical modality), about counterfactual conditionals, about objective chance or propensity (as a generalization of objective modality), and about physical or natural necessity as embodied in *natural laws*. Overcoming this obstacle calls for a revolutionary rethinking of our standard picture of causation. This standard picture I call the *horizontal* or *externalist* model of causation. The alternative I am proposing is the thesis of *causal internalism*, which countenances the reality of *vertical causation*.

On the standard, horizontal model, causes and effects are, exclusively, physical, spatio-temporally local states and occurrences. The causal nexus, whether

it consists in a kind of necessary, stochastic, or nomic connection, stands outside of both the cause and the effect. This is why I call it *causal externalism*: the causal nexus is wholly external to both the cause and the effect. The horizontal/externalist model can account for our knowledge of occurrent properties realized in spatiotemporal locations, but it leaves the entire realm of modality causally, and, therefore, cognitively and epistemically, inaccessible.

My alternative proposal is that we consider the modal (or nomic or stochastic) facts that tie the cause to the effect to be *internal* to the cause or to the effect. In this paper, I will focus on the moments or data of modal reality that are internal to the *cause*. Depending on the details of one's account of causation, causes necessitate or probabilify or possibilify their effects. On an internalist model, the fact that a given cause necessitates its effect is itself an integral part of the total cause, not something that stands outside or above the cause-effect pair. Consequently, modal facts are every bit as causally efficacious as are occurrent physical facts, and so there is no barrier to providing a unified, causal theory of all of human thought and knowledge. For instance, we can think about and gain knowledge of natural laws by virtue of the fact that each of these laws enter into some, but not all, causal connections. When we observe a regularity (like the elliptical orbits of the planets) that is really caused by a particular nomic fact (like the law of gravitation), then our observations provide us with intentional and epistemic contact with that nomic fact.

In my forthcoming book, *Realism Regained*, [18] I develop three models of causation: deterministic, indeterministic and probabilistic. In the deterministic model, causes are strictly sufficient for (i.e., necessitate) their effects. In the indeterministic model, causes are *para-sufficient* or defeasibly sufficient for their effects, a relation I model by means of recent work on defeasible or nonmonotonic logics. Finally, in the probabilistic model, there is an irreducibly probabilistic or stochastic relationship between the cause and the effect. In the book, I give number of reasons for preferring the indeterministic and probabilistic models, but for simplicity's sake, I will present only the deterministic model in this paper.

2 Basic Ontology

In this section, I will introduce the model structures to be taken as formal representations of real possibilities concerning causation. These structures incorporate two kinds of individuals: situation-tokens and situation-types [Barwise and Perry (1983), Barwise and Etchemendy (1987), Barwise (1989)]. Actual situation-tokens are to be thought of as real, concrete parts of the world, analogous to Davidsonian events [Davidson (1980)]. Merely possible situation-tokens are abstract objects, constructible from actual tokens and types, representing possible but unrealized actualities. Each token carries a certain amount of information or fact about the world: these units of fact are represented as situation-types.

2.0.1 Classification Systems

A classification system consists of a set of tokens, a set of types, and a binary relation on the two sets (the classification relation). (These structures have been independently discovered many times over. Birkhoff (1940)[4] called them “polarities”, and Hardegree (1982)[15] called them “contexts”. They were also invented by the German mathematician Wille, whose work is discussed in Davey and Priestley (1990).[7]) For my purposes, the set of tokens will be a set of situation-tokens, the set of types situation-types, and the classification relation \models .

I will assume that the set of types is closed under the Boolean operators ‘ \neg ’ and ‘ \vee ’. The classification relation \models can be constrained to satisfy such logical principles as De Morgan’s laws, distribution, associativity and commutativity of \vee , and the laws of double negation. In addition, if $s \models \phi$, then $s \not\models \neg\phi$, but not necessarily vice versa. (So, ‘ \neg ’ on types can represent weak or internal negation.)

2.1 Models

In the most general case, a model would contain a set of tokens, Sit , together with a function \mathcal{S} that assigns a classification system to each situation. In addition, we need two partial orderings on situation-tokens, \sqsubseteq and \prec . The first represents the part-whole relationship of standard mereology. The second, a strict partial well-ordering, represents the relation of causal precedence. In representing causation, we can look at a simpler, special case, one in which all of the classification systems share the same set Typ of types and the same classification relation \models . In this special case, the function \mathcal{S} assigns to each situation-token s a subset of Sit . The tokens in $\mathcal{S}(s)$ are those that are possible alternatives, from the perspective of s , that is, the set $\mathcal{S}(s)$ represents the modal facts about the world as they are supported by s .

$\mathcal{W}(s)$ shall be the \sqsubseteq -maximal members of $\mathcal{S}(s)$. These situations represent possible “worlds”, from the perspective of s .

Consequently, a *standard, deterministic model* \mathcal{M} consists of an n -tuple, $\langle Sit, Typ, \models, \mathcal{S}, \sqsubseteq, \prec \rangle$, where:

- Sit is a nonempty set, the set of situation-tokens.
- Typ is a nonempty set of situation-types, closed under the Boolean operators \vee and \neg .
- \models is a binary relation on $Sit \times Typ$.
- \mathcal{S} is a function from Sit to the powerset of Sit .
- \sqsubseteq is a partial ordering of Sit (the part/whole relation of mereology). There is a set \mathcal{W} of maximal situations (worlds). Every situation is extended by some world.
- \prec is a partial ordering of Sit .

2.2 The Causal Priority Relation

The causal priority relation \prec cannot be identified simply with causation. Instead, it represents a necessary pre-condition for causation. In fact, under the assumption of determinism, these three causal notions are interdefinable:

- \prec , the relation of causal priority.
- \triangleright , the relation of being a total cause of.
- \rightsquigarrow , the relation of being an essential part of a total cause of, Mackie's INUS condition: "an Insufficient but Necessary part of an Unnecessary but Sufficient condition" [Mackie (1965)].

In addition, we can say that two tokens are *coincident* if they share exactly the same causal antecedents.

Definition 2.1 (Coincidence) $s \approx s' \Leftrightarrow \forall s'' \in \text{Sit}(s'' \prec s \leftrightarrow s'' \prec s')$

I will assume that the causal priority relation is transitive, irreflexive, and well-founded. A token s is immediately prior to s' just in case s is prior to s' , and there are no intermediate tokens.

$$(s \prec_0 s') \leftrightarrow_{df} (s \prec s') \& \neg \exists s'' (s \prec s'' \& s'' \prec s')$$

The three causal relations \prec , \rightsquigarrow and \triangleright are such that any two of these can be defined in terms of the third, together with the mereological part-whole relation \sqsubseteq :

- $s \prec s'$ iff there is a w such that s is a part of the mereological sum of INUS causes of s' in w (or equivalently, iff s is a part of the mereological sum of minimal total causes of s' in w).
- $w \models s \rightsquigarrow s'$ iff s is part of some minimal total cause of s' in w .
- $w \models s \triangleright s'$ iff $s \prec s'$, $s, s' \sqsubseteq w$, and s necessitates s' .

More or less arbitrarily, I will take the causal priority relation to be primitive and define the other two in terms of it. In this case, the first condition above imposes a minimality requirement on the extension of \prec in a model. In *Realism Regained*, I argue that causal priority need not be taken as a primitive relation when using the indeterministic or probabilistic models of causation. Instead, in those cases, causal priority can be identified with *asymmetric token necessitation*: a token a is causally prior to token b just in case the two tokens do not overlap, the existence of b necessitates the existence of a , and not vice versa. Obviously, this definition is unavailable in the deterministic model, since that model includes the thesis that a cause necessitates its effect.

3 Partial Propositional Logics

It is conventional in philosophical logic to refer to both three- and four-valued logics as “partial logics”. A three-valued logic recognizes the values true, false and neither (undetermined). Four-valued logic adds the possible value of both true and false. Three-valued logics are useful in representing partially undefined or ontologically incomplete situations. Four-valued logics enable us to deal both with incomplete and with logically impossible situations.

For most purposes, three-valued situation theory will suffice, since all actual and possible situation-tokens bear one of three relations to each situation-type (verify, falsify or neither). However, there are two reasons for being interested in four-valued logic. First, there are concerns of symmetry and elegance. The four values form a kind of lattice, and in many cases, the logic and semantics of four-valued systems are simpler than those of the corresponding three-valued systems. Second, I am interested in situations which are partial with respect to information about logical necessity. Such situations do not recognize the impossibility of certain logically impossible situations. In order to model such logical impossibility, it is convenient to make use of the fiction of logically impossible (four-valued) tokens. Logically partial situation-tokens will enable us to recognize the causal efficacy of logical necessities, which in turn will make possible a causal theory of logical reference and knowledge.

Here are the three-valued (strong Kleene) truth tables for negation, disjunction, and conjunction.

$\neg p$	T	p	F	U
$\neg p$	F	T	U	

$(p \vee q)$		p	T	F	U
q	T	T	T	T	T
	F	T	F	F	U
	U	T	U	U	U

$(p \& q)$		p	T	F	U
q	T	T	F	F	U
	F	F	F	F	F
	U	U	F	F	U

The corresponding tables for four-valued logic (the Dunn tables) [10] are as follows:

$\neg p$	T	p	F	U	B
$\neg p$	F	T	U	B	

$(p \vee q)$		p		
	T	F	U	B
q	T	T	T	T
	F	T	F	B
	U	T	U	T
	B	T	B	B

$(p \& q)$		p		
	T	F	U	B
q	T	T	F	B
	F	F	F	F
	U	U	F	F
	B	B	F	B

The unifying idea behind both the strong Kleene and the Dunn tables is simply this: one computes truth-values as one would in classical semantics, except that one separates the determination of truth and of falsehood:

1. $\neg\phi$ is (at least) true if ϕ is.
2. $\neg\phi$ is (at least) false if ϕ is.
3. $\phi\&\psi$ is at least true if both ϕ and ψ are.
4. $\phi\&\psi$ is at least false if either ϕ or ψ are.
5. $\phi\vee\psi$ is at least true if either ϕ or ψ are.
6. $\phi\vee\psi$ is at least false if both ϕ and ψ are.

By ‘at least true’, I mean either T or B (either true only or both true and false), and by ‘at least false’ I mean either F or B (either false only or both true and false). A proposition that receives no truth value from these principle retains the classification U .

The use, within semantic models, of impossible situation-tokens is no more troubling, ontologically speaking, than the use of possible-but-not-actual tokens. The only real tokens are actual tokens. Non-actual tokens, whether possible or not, are useful merely fictions. Modality (possibility, contingency, necessity) pertains primarily to actual situation-types. A paradigmatic modal fact is something like: type ϕ is possibly instantiated. Since the work of Kripke and Kanger, it is widely recognized that it is useful, in representing the semantics and logic of modal facts, to construct models containing indices that stand for merely possible worlds. Each possible world represents the compossibility (from the point of view of certain worlds) of a set of types. Similarly, in four-valued modal logic, I will use impossible situation-tokens to represent lack of impossibility (from the point of view of logically partial situations) of the co-exemplification of certain types.

In partial logic, there are no logically true propositions: no proposition is true in every three- or four-valued interpretation. We do, however, have non-trivial

logical implication. In fact, there are a variety of relations that are species of logical implication or consequence in partial logic. In three-valued logic, there are three notions of logical consequence that seem most natural: verification validity, falsification validity, and double-barrelled validity. A set Γ verifiably entails a set Δ just in case every interpretation that verifies every member of Γ verifies some member of Δ . A set Γ falsifiably entails a set Δ just in case every interpretation that falsifies every member of Γ falsifies some member of Δ . The relation of double-barrelled implication (first suggested by Blamey[5]) holds between a set Γ and Δ just in case Γ both verifiably and falsifiably implies Δ . Whenever I talk about implication in three-valued logic, I will mean double-barrelled implication, since this comes closest in many ways to the classical case.

In four-valued logic, the situation is much simpler, in that verification implication, falsification implication and double-barrelled implication all coincide.[23, page 77]

Muskens[23] has proved that the following system of rules (rL^{+*}) is complete for double-barrelled implication in three-valued propositional logic:

- (R1) $\neg\neg\phi \vdash \phi$
- (R2) $\neg(\phi \& \psi) \vdash \neg\phi \vee \neg\psi$
- (R3) $\neg(\phi \vee \psi) \vdash \neg\phi \& \neg\psi$
- (R4) $\phi \& \psi \vdash \phi$ and $\phi \& \psi \vdash \psi$
- (R5) $\phi \vdash \phi \vee \psi$ and $\psi \vdash \phi \vee \psi$
- (R6) If $\phi, \rho \vdash \chi$, and $\psi, \rho \vdash \chi$, then $(\phi \vee \psi), \rho \vdash \chi$.
- (R7) If $\chi \vdash \rho, \phi$, and $\chi \vdash \psi, \rho$, then $\chi \vdash (\phi \& \psi), \rho$.
- (R8*) $\phi \& \neg\phi \vdash \psi \vee \neg\psi$
- (R9) If $\phi \vdash \psi$ and $\psi \vdash \chi$, then $\phi \vdash \chi$.
- (R10) $\Gamma \vdash \Delta$ if and only if there are nonempty sets $\{\alpha_1, \dots, \alpha_m\} \subseteq \Gamma$ and $\{\beta_1, \dots, \beta_n\} \subseteq \Delta$ such that $\alpha_1 \& \dots \& \alpha_m \vdash \beta_1 \vee \dots \vee \beta_n$.

For four-valued logic, Muskens demonstrates that the system rL is sound and complete, where rL is $rL^{+*} - (R8^*)$.

4 Partial Modal Logics

I take it as obvious that there is some important connection between causation and modality. As Hume famously observed, causation seems to involve some kind of “necessary connection”. Consequently, I need to make use of a partial version of modal logic. Fortunately, the groundwork for partial modal logic has

already been laid by Thijsse[30] and Muskens[23]. In this work, I will follow Muskens very closely.

A model in partial model logic shall consist of a quintuple $\langle Sit, R^\uparrow, R^\downarrow, I, \sqsubseteq \rangle$. Sit is the set of situation-tokens, which are essentially partial (incomplete or overdetermined) worlds. The part-whole relation \sqsubseteq is a partial ordering on the class Sit . The interpretation function assigns truth values (true, false, neither, or both) to atomic symbols, representing persistent situation-types. Whenever $s \sqsubseteq s'$, I will require that $I(\phi, s')$ be an enrichment of $I(\phi, s)$, for every atomic symbol ϕ . (Every other value is an enrichment of the value undefined, and the value both is an enrichment of the two classical values.)

The relations R^\uparrow and R^\downarrow are binary relations on Sit . These are the outer and inner accessibility relations. If we have $sR^\uparrow s'$, then we fail to falsify the accessibility of s' from s : the model treats the accessibility of s' from s as either definitely established or undefined. Dually, if we have $sR^\downarrow s'$, then we definitely verify the accessibility of s' from s . This gives us four possible values for the accessibility of s' from s :

- The relation is undefined: we have $sR^\uparrow s'$, but not $sR^\downarrow s'$.
- The relation is verified only: we have both $sR^\uparrow s'$ and $sR^\downarrow s'$.
- The relation is falsified only: we have neither $sR^\uparrow s'$ nor $sR^\downarrow s'$.
- The relation is both falsified and verified: we have $sR^\downarrow s'$, but not $sR^\uparrow s'$.

Whenever a situation-token s is logically possible, we have both that $I(\phi)(s)$ is true, false or undefined for every ϕ (never both true and false), and the image $R^\downarrow[\{s\}]$ is a subset of the image $R^\uparrow[\{s\}]$ (no situation is both definitely accessible to s and definitely not accessible to s).

I will require that modal facts be persistent. So, if $s \sqsubseteq s'$, then $R^\uparrow[\{s'\}] \subseteq R^\uparrow[\{s\}]$, and $R^\downarrow[\{s'\}] \subseteq R^\downarrow[\{s\}]$. In other words, as we move from a smaller to a larger situation-token, the set of definitely inaccessible tokens monotonically increases, as does the set of definitely accessible tokens.

The truth and falsity definitions for the modal operators are quite simple:

- $\mathcal{M}, s \models \Box\phi \Leftrightarrow \neg\exists x \in R^\uparrow[\{s\}]\mathcal{M}, x \models \neg\phi$.
- $\mathcal{M}, s \models \neg\Box\phi \Leftrightarrow \exists x \in R^\downarrow[\{s\}]\mathcal{M}, x \models \neg\phi$.
- $\mathcal{M}, s \models \Diamond\phi \Leftrightarrow \exists x \in R^\downarrow[\{s\}]\mathcal{M}, x \models \phi$.
- $\mathcal{M}, s \models \neg\Diamond\phi \Leftrightarrow \neg\exists x \in R^\uparrow[\{s\}]\mathcal{M}, x \models \phi$.

The formula $\Box\phi$ (the necessity of ϕ) is true in a model at a situation s just in case ϕ is never falsified by any token in the outer accessibility set of s (the set of tokens that are not definitely inaccessible). This formula is false in a model at s if it is falsified by some situation-token in the inner accessibility set of s (the tokens that are definitely accessible). The possibility operator \Diamond is defined as the dual of \Box .

In these definitions, I deviate from the pattern of both Thijsse and Muskens, since for my purposes, it is essential that I make all modal facts persistent (with respect to moving up the part-whole ordering \sqsubseteq), which the Thijsse-Muskens truth definitions fail to do. Nonetheless, it is easy to demonstrate that the logic of double-barrelled consequence in the three-valued case is characterized by the Thijsse system M^K , and the four-valued logic is characterized by the system M^{--} , one of Thijsse's systems [30, page 104].

The system M^K consists of rL^{+*} plus the following rules:

- (R11) $\diamond\neg\phi \vdash \neg\Box\phi$
- (R12) $\Box\neg\phi \vdash \neg\diamond\phi$
- (R13) $\diamond(\phi \vee \psi) \vdash \diamond\phi \vee \diamond\psi$
- (R14) If $\phi \vdash \psi$, then $\diamond\phi \vdash \diamond\psi$.
- (R15) $\diamond(\phi \& \neg\phi) \vdash \Box(\psi \vee \neg\psi)$
- (K) $\Box(\phi \rightarrow \psi), \Box\phi \vdash \Box\psi$
- (KNec) If ϕ is a theorem of the classical (two-valued) system K , then $\vdash \Box\phi$.

The system M^K agrees with the classical modal logic K with respect to all theorems that are preceded by a box, since it is impossible to find coherent models that falsify any classical validity, and the truth definition of $\Box\phi$ guarantees its truth whenever ϕ cannot be falsified.

The system M^{--} , which characterizes four-valued modal logic, consists of rL plus rules (R11) through (R14), plus two additional rules, (R16) and (R17):

- (R16) $\Box\phi \& \Box\psi \vdash \Box(\phi \& \psi)$
- (R17) If $\phi \vdash \psi$, then $\Box\phi \vdash \Box\psi$.

The completeness of these two systems for their respective sets of models can be demonstrated quite easily by means of the standard construction of canonical models. In the case of three-valued (or coherent) modal logic, the set *Sit* in the canonical model consists of the set of consistent, saturated theories of the logic M^K . A theory of a modal system is a set of sentences that is closed under the rules of the system. A theory is consistent if it does not contain both ϕ and $\neg\phi$, for any formula ϕ . A theory is saturated if it contains either ϕ or ψ whenever it contains the disjunction $\phi \vee \psi$. The interpretation function I for the canonical model is defined thus: for each atomic formula ϕ , if $\phi \in \Gamma$, then $I(\phi)(\Gamma) = T$, if $\neg\phi \in \Gamma$, then $I(\phi)(\Gamma) = F$, and otherwise $I(\phi)(\Gamma) = U$. Two theories Γ and Δ are in the part-whole relation \sqsubseteq just in case Γ is a subset of Δ . Finally, we need to define the two partial accessibility relations R^\uparrow and R^\downarrow for the canonical model:

$$\Gamma R^\uparrow \Delta \Leftrightarrow \forall \psi (' \Box \psi ' \in \Gamma \rightarrow ' \neg \psi ' \notin \Delta)$$

$$\Gamma R^\downarrow \Delta \Leftrightarrow \forall \psi (' \psi ' \in \Delta \rightarrow ' \Diamond \psi ' \in \Gamma)$$

In the case of four-valued (general) partial modal logic, the canonical model is constructed in exactly the same way, except that the class *Sit* is the set of all saturated theories of M^{--} , whether or not they are consistent.

For these completeness proofs, we need a partial version of Lindenbaum's Lemma:

Lemma 1 (Lindenbaum's Lemma for Partial Modal Logic) *If $\Gamma \not\vdash \Delta$, then there is a saturated theory Γ' such that $\Gamma \subseteq \Gamma'$ and $\Gamma' \cup \Delta = \emptyset$.*

This partial Lindenbaum's lemma can be proved by a slight modification of the usual proof.[23]

We then prove by induction a canonical model theorem, showing that for every situation s in the canonical model (that is, for every saturated theory), and every formula ϕ of the language, $\phi \in s$ if and only if $\mathcal{M}_{Can}, s \models \phi$, and $\neg\phi \in s$ if and only if $\mathcal{M}_{Can}, s \models \neg\phi$. This proof follows the usual one in the atomic cases and in the cases corresponding to the propositional connectives. In the case of the modal connectives, we must make use of the partial-logic Lindenbaum's lemma.

4.1 Reflexive Models

If we require, as seems natural, that the accessibility relations be reflexive, then we can strengthen the two systems M^K and M^{--} . In the case of three-valued modal logic, the class of reflexive models can be characterized by adding two new rules:

$$\text{(Refl1)} \quad (\phi \& \Box \neg\phi) \vdash \neg\phi$$

$$\text{(Refl2)} \quad \phi \vdash (\Diamond\phi \vee \neg\phi)$$

In the four-valued case, I do not believe that we can characterize the class of models in which R^\downarrow is reflexive.

One solution to this latter problem is to introduce a hybrid logic. In models of this logic, there is a single, designated situation g (the actual situation, intuitively). We can require that g be coherent, in the sense that $I(g)$ assigns only T , F or U to all atomic formulas, and $R^\downarrow[\{g\}] \subseteq R^\uparrow[\{g\}]$, and in addition, we could require that every situation in $R^\downarrow[\{g\}]$ (the situations that are definitely accessible to g) be similarly coherent. Other situations in the model, however, could be logically incoherent, requiring the use of the four-valued truth functions. We could define logical consequence for this system by reference to these distinguished worlds g : Γ entails Δ if and only if:

1. every model \mathcal{M} such that \mathcal{M}, g verifies every member of Γ is such that $\mathcal{M}, g_{\mathcal{M}}$ verifies some member of Δ , and
2. every model \mathcal{M} such that \mathcal{M}, g falsifies every member of Δ is such that $\mathcal{M}, g_{\mathcal{M}}$ falsifies some member of Γ .

The logic for this system, M^H , would consist of rules (R1) through (R17), but would lack rules (K) and (KNec). In addition, we could characterize the class of reflexive models by using rules (Refl1) and (Refl2).

If we restrict our attention to possible *worlds*, situation tokens that are coherent and complete, then we would return to the classical two-valued modal logics, like T, S4 or S5. My intention is not to argue that partial modal logic should *replace* classical modal logic. We still need classical modal logic to characterize a certain kind of validity. For instance, the inference from $\Box\phi$ to ϕ is not *locally valid*, since there are many situation tokens that verify the first but not the second. However, this same inference is *globally valid*, since any token that verifies the first is embedded in a possible world that somewhere verifies the second. The corresponding T axiom, $(\Box\phi \rightarrow \phi)$, is not verified at every situation token, but only because many tokens contain only partial information about modality. As the modal information supported by a token is enriched, we approach classical modal logic at the limit. Partial modal logic is important in representing facts about causal connections between partial tokens, as we shall see in the following sections.

5 Partiality and Quantificational Logic

Partial modal logic, as important as it is, is not in themselves sufficient for formulating an adequate model of causation. In addition to the representation of necessity and objective probability, we must also be able to talk about the situations that are causing and being caused, and we must be able to represent some situations as parts of others. Consequently, in this section I will develop a partial quantificational logic.

This quantificational logic will be different from standard quantified modal logics in that the individuals being named and quantified over will be indices (situations and worlds) and not ordinary substances (like people and organisms). One might think that quantification over situations makes the modal operators redundant, since we could define necessity by simply using a universal quantifier. However, replacing modality with explicit quantification over situations would eliminate a critical element implicit in the use of modal operators: the *indexicality* of modal properties. We could re-introduce this element of indexicality by adding a special, indexical constant to our language, something that intuitively picks out “this situation”. The necessity of ϕ could then be defined as ϕ ’s holding in every situation accessible from *this situation*. However, this fix would introduce another problem: many formulas involving the constant *this situation* would be non-persistent. We could have the formula “ ψ does not hold

in *this situation*” holding in s but not holding in a strictly larger situation s' that contains s .

Thus, I will use a language that contains both modal operators (with their implicit indexicality) and terms and variables that stand for situations (and do so non-indexically). In addition to situation constants, variables, and quantifiers, I will add two new kinds of atomic formulas (t and t' are situation constants and ϕ is any formula):

- $(t \equiv \phi)$
- At

We can define the part-whole relation \sqsubseteq by means of these elements:

$$(t \sqsubseteq t') =_{def} (t' \equiv At)$$

In turn, the identity predicate can be defined in terms of \sqsubseteq : $(t = t') =_{def} (t \sqsubseteq t' \ \& \ t' \sqsubseteq t)$.

The first new kind of atomic formula is an object language counterpart to the verification relation between situation tokens and types. For simplicity’s sake, I will assume that the logic of these two kinds of formulas is entirely classical (bivalent). I will assume that if one situation is part of another, or if one situation verifies a formula, then these facts are supported in every situation in the model. For some purposes it might be useful to make situations partial in their mereological or classificatory information (for example, this might be very important in modeling certain propositional attitudes), but I have not found this additional flexibility necessary in dealing with the concept of causation. Thus, the truth and falsity conditions for these formulas are the following (where $\|t\|$ represents the designatum of constant t in the model \mathcal{M}):

- $\mathcal{M}, s \models (t \sqsubseteq t') \Leftrightarrow \|t\| \sqsubseteq \|t'\|$
- $\mathcal{M}, s \models \neg(t \sqsubseteq t') \Leftrightarrow \neg(\|t\| \sqsubseteq \|t'\|)$
- $\mathcal{M}, s \models (t \equiv \phi) \Leftrightarrow \mathcal{M}, \|t\| \models \phi$
- $\mathcal{M}, s \models \neg(t \equiv \phi) \Leftrightarrow \mathcal{M}, \|t\| \not\models \phi$

The atomic formula At means that the situation t is definitely part of the actual world (from the perspective of the indexed situation). Such a formula is verified by a situation s just in case the situation $\|t\|$ is a part (proper or improper) of s . A model structure for the language must include a binary relation A^- to provide falsity conditions for the A predicate.

- $\mathcal{M}, s \models At \Leftrightarrow \|t\| \sqsubseteq s$
- $\mathcal{M}, s \models \neg At \Leftrightarrow \langle s, \|t\| \rangle \in A^-$

To ensure that the logic of situations is axiomatizable, it is essential that we place certain conditions on the relation A^- . First, there is a fixed-point condition: whenever two situations verify contradictory formulas of any kind (including formulas involving A itself), the ordered pair of the two situations must belong to A^- . Second, if a situation s does not support $\neg\phi$, then there must exist a token s' such that s' supports ϕ and $\langle s, s' \rangle$ does not belong to the relation A^- . Third, if there is no situation accessible to situation s (either by the inner relation R^\downarrow or by the outer relations R^\uparrow) that extends both s and s' , then the pair $\langle s, s' \rangle$ belongs to A^- . Finally, the relation A^- must be mereologically persistent.

1. (Fixed point condition) If $\mathcal{M}, s \models \phi$ and $\mathcal{M}, s' \models \neg\phi$, then $\langle s, s' \rangle \in A^-$.
2. If $\mathcal{M}, s \not\models \neg\phi$, then there exists an s' such that $\mathcal{M}, s' \models \phi$ and $\langle s, s' \rangle \notin A^-$.
3. (Modal condition) If $\neg\exists x(x \in R^\uparrow[\{s\}] \cup R^\downarrow[\{s\}] \& s \sqsubseteq x \& s' \sqsubseteq x)$, then $\langle s, s' \rangle \in A^-$.
4. If $\langle x, y \rangle \in A^-$, and $x \sqsubseteq z$ and $y \sqsubseteq w$, then $\langle z, w \rangle \in A^-$.

In the case of the quantifiers, the truth and falsity conditions reflect the usual extension of the Dunn truth tables.

- $\mathcal{M}, s \models \forall x\phi \Leftrightarrow$ for every situation s' in Sit , $\mathcal{M}_{s'}, s \models \phi[t/x]$, where $\mathcal{M}_{s'}$ is a model that differs from \mathcal{M} only in assigning s' to a new constant t (a constant not occurring in ϕ).
- $\mathcal{M}, s \models \neg\forall x\phi \Leftrightarrow$ for some situation s' in Sit , $\mathcal{M}_{s'}, s \models \neg\phi[t/x]$, where $\mathcal{M}_{s'}$ is a model that differs from \mathcal{M} only in assigning s' to a new constant t (a constant not occurring in ϕ).

The logic that corresponds to this semantical system will be called \mathcal{L}_{Sit} . In addition to the rules of the system rL for partial propositional logic, M^{--} for partial model logic, and rC for partial conditional logic, we need to add, first, the following quantifier and identity rules:

- (Q1) $\forall x\phi \vdash \phi[t/x]$
- (Q2) $\phi[t/x] \vdash \exists x\phi$
- (Q3) If $\Gamma \vdash \phi[t/x]$ and t does not occur in Γ or ϕ , then $\Gamma \vdash \forall x\phi$.
- (Q4) If $\Gamma, \phi[t/x] \vdash \Delta$, and t does not occur in Γ, ϕ or Δ , then $\Gamma, \exists x\phi \vdash \Delta$.
- (Q5) $\vdash \forall x\forall y(x = y \leftrightarrow ((x \equiv Ay) \& (y \equiv Ax)))$
- (Q6) $\vdash \forall x(x \equiv Ax)$
- (Q7) $\phi \vdash \forall x(x = t \rightarrow \phi[x//t])$, where $\phi[x//t]$ is the result of replacing one or more occurrences of t in ϕ with x .

- (Q8) $\neg\forall x\phi \vdash \exists x\neg\phi$
- (Q9) $\neg\exists x\phi \vdash \forall x\neg\phi$

In order to capture the classicality of atomic formulas involving $|\equiv$, we must add the following two rules:

- (Q8) If Γ classically entails Δ , and the atomic formulas in Γ and Δ involve only $|\equiv$, then $\Gamma \vdash \Delta$.
- (Q9) $(t|\equiv \phi), \neg(t|\equiv \phi) \vdash \psi$

Rule (Q10) connects impossibility with the support of non-actuality. Since all the formulas of our language are persistent (with respect to the part-whole ordering), we must add rules that ensure that if a situation t is part of the current index, and t supports formula ϕ , then the current index also supports ϕ . Also, if the current index supports ϕ , and t supports $\neg\phi$, then the current index supports $\neg At$. Thirdly, every situation verifies the formula that it is actual. Rules (Q11) and (Q12) guarantee that actuality has the right kind of fixed-point character.

- (Q10) $At, \Box\neg(At \& At') \vdash \neg At' \vee \Diamond(At \& \neg At')$
- (Q11) $\phi \vdash \exists x(Ax \& (x|\equiv \phi))$
- (Q12) $\neg\phi \vdash \forall x((x|\equiv phi) \rightarrow \neg Ax)$

The rules governing the *support* relation $|\equiv$ ensure that the formulas supported by a situation form a saturated theory.

- (Q13) $\vdash \forall x((x|\equiv \forall y\phi) \leftrightarrow \forall y(x|\equiv \phi))$
- (Q14) $\vdash \forall x((x|\equiv (\phi \& \psi)) \leftrightarrow (x|\equiv \phi) \& (x|\equiv \psi))$
- (Q15) $\vdash \forall x((x|\equiv \exists y\phi) \leftrightarrow \exists y(x|\equiv \phi))$
- (Q16) $\vdash \forall x((x|\equiv (\phi \vee \psi)) \leftrightarrow (x|\equiv \phi \vee x|\equiv \psi))$
- (Q17) If $\phi \vdash \psi$, then $(t|\equiv \phi) \vdash (t|\equiv \psi)$.

I will assume that all mereological and classificatory facts are supported by all tokens, and that they hold of necessity:

- (Q18) $\vdash \forall x\forall y((x|\equiv (y|\equiv \phi)) \leftrightarrow (y|\equiv \phi))$
- (Q19) $\vdash \forall x\forall y((x|\equiv \neg(y|\equiv \phi)) \leftrightarrow \neg(y|\equiv \phi))$
- (Q20) $\Diamond\phi \vdash \phi \vdash \Box\phi$, where ϕ contains only $|\equiv$ atoms.

Finally, I will stipulate that the domains of quantification associated with all situations are the same. In each case, we will be quantifying over all situations, actual and non-actual, possible and impossible. We can of course express the actuality of a situation by the formula At and its possibility by $\Diamond At$. Since the domains of quantification are constant, we can add two rules corresponding to the Barcan and converse Barcan formulas of standard quantified modal logic.

- (Q21) $\forall x \Box \phi \vdash \Box \forall x \phi$
- (Q22) $\forall x \Diamond \phi \vdash \Diamond \forall x \phi$

In a canonical model for the logic \mathcal{L}_{Sit} , the set of situations consists of a certain set of supersaturated theories. A theory Γ is *supersaturated* if and only if it meets the three conditions:

1. If $\forall x \phi \notin \Gamma$, then for some t , $\phi[t/x] \notin \Gamma$.
2. If $\exists x \phi \in \Gamma$, then for some t , $\phi[t/x] \in \Gamma$.
3. If $(\phi \vee \psi) \in \Gamma$, then either $\phi \in \Gamma$ or $\psi \in \Gamma$.

6 Constraints and Causation at the Token Level

On the deterministic conception, a token s causes token s' if three conditions are met: (i) token s is actual, (ii) s is causally prior to s' , and (iii) the actuality of s constrains the world to contain s' as well, in other words, s objectively necessitates s' . This notion of constraint or necessitation can be defined thus.

Definition 6.1 (Token-to-token Constraint) $(s_1 \vdash s_2) =_{def} \Box(A s_1 \rightarrow A s_2)$

This definition can also be generalize to a relation between tokens and sets of tokens (assuming that our language has been enriched by some means of referring to sets of tokens). Token s constrains the actuality of set B just in case it necessitates the actuality of some member of B :

$$(s_1 \vdash B) =_{def} \Box(A s_1 \rightarrow \exists x \in B A x)$$

Definition 6.2 (Token Causation under Determinism)

$$(s_1 \triangleright s_2) =_{def} A s_1 \ \& \ (s_1 \prec_0 s_2) \ \& \ (s_1 | \equiv (s_1 \vdash s_2))$$

This definition of token causation can also be generalized to a relation between tokens and sets of tokens:

$$(s_1 \triangleright B) =_{def} A s_1 \ \& \ \forall x \in B (s_1 \prec_0 x) \ \& \ (s_1 | \equiv (s_1 \vdash B))$$

The token-to-token constraint relation is one of strict necessitation: every world containing the first situation must also contain the second. Token determinism consists of two theses: causes must be actual, and causes constrain their effects to be actual as well. Given the definition of constraint, it follows that if s is a world, and $s \models (s_1 \triangleright s_2)$, then both s_1 and s_2 must be actual in (parts of) s .

Notice that, in order for $s_1 \triangleright s_2$ to be true in situation s , it is necessary that s_1 , and not just s , support the constraint $s_1 \vdash s_2$. **This requires that a sufficient number of relevant modal facts are incorporated into the total cause.** It is this *causal internalism* that makes possible the representation of the kind of *vertical causation* needed for a causal theory of modal knowledge and cognition.

6.1 Type/type Constraints

We can also define causal constraints between situation-types. In order to do so, I must first define a causal succession relation between tokens, abbreviated as sNs' .

Definition 6.3 (Causal Succession)

$$\begin{aligned} \forall x \forall y (xNy =_{def} \\ \forall z (z \sqsubseteq y \leftrightarrow (Az \& (x \prec_0 z)))) \end{aligned}$$

Definition 6.4 (Causal constraint on types)

$$\begin{aligned} (\phi \sim \psi) =_{def} \Box \forall x ((Ax \& (x \equiv \phi) \rightarrow \\ \exists y (xNy \& (y \equiv \psi))) \end{aligned}$$

A causally informed constraint from ϕ to ψ entails that every ϕ -situation must be immediately followed by a ψ -situation.

Type constraints give rise to a distinctive form of modal logic. Since we are working with partial, three- or four-valued worlds, substitution into modal contexts is permissible only if the relevant types are strong-Kleene or Dunn equivalent, not just classically equivalent. For example, ϕ and $((\phi \& \psi) \vee (\phi \& \neg \psi))$ are classically, but not strong-Kleene or Dunn, equivalent. This hyperintensionality of causal contexts is vital to their use in explicating teleological and representational properties.

7 Defining Causal Explanation (or Fact/Fact Causation)

A *causal explanation* is a relation between one token-type pair and another token-type pair. A pair $\langle s, \phi \rangle$ causally explains a pair $\langle s', \psi \rangle$ just in case s caused s' , and s 's being of type ϕ explains why its effect had to be of type

ψ . This corresponds closely to Terence Horgan's notion of *qua*causation: s qua ϕ causes s' qua ψ . [16][20] I will use this notion to define the causal powers of a property, and it can also be used to define the causal relevance of a particular instantiation of a property to other facts.

My definition of causal explanation is intended to capture only the metaphysical core of our ordinary notion of 'explanation'. As many have observed, there are a host of pragmatic factors that enter into making something a good or apt explanation. Explanation in this full, pragmatic sense is typically contrastive: we explain why something is ϕ *as opposed to* ψ . Moreover, explanation depends on the knowledge and interests of the audience: we do not typically cite something that everyone knows was present, such as including the presence of oxygen in the atmosphere as part of the explanation of a house fire. However, I am aiming at a characterization of an interest-independent, non-pragmatic explanatory relation, one that constitutes a necessary condition of something's being a correct explanation. This relation could also be thought of as a relation of objective causation between *facts*, where facts are identified with pairs consisting of an actual situation-token and a type that it supports.

Definition 7.1 (Causal Explanation (Fact/Fact Causation))

$$\begin{aligned} ((s_1 : \phi) \triangleright (s_2 : \psi)) =_{def} \\ As_1 \ \& \ s_1 N s_2 \ \& \ (\phi | \sim \psi) \ \& \ (s_1 | \equiv \phi) \end{aligned}$$

Causal explanation is veridical in both terms: both s_1 and s_2 must be parts of s , s_1 must be of type ϕ , and s_2 must be of type ψ . It is also irreflexive: no token-type pair explains itself. The transitive closure of the explanation relation would also be irreflexive, so explanatory loops are excluded. If the transitive closure of the \prec relation is a partial well-ordering, there cannot be any explanatory infinite regresses.

As I just said, causal explanation under the deterministic conception is provably sound: if there exists an explanation of s 's being ψ , then s really is ψ . In contrast, there is no necessity that explanation be complete: that there be an explanation for every type characterizing every causally consequent situation. Thus, the deterministic conception of causation does not by itself guarantee that type determinism is true. We can consider the completeness of causal explanation as an optional hypothesis.

Theorem 1 (Soundness of Causal Explanation)

$$((s_1 : \phi) \triangleright (s_2 : \psi)) \Rightarrow As_2 \ \& \ (s_2 | \equiv \psi)$$

Proof A trivial consequence of the definitions.

7.1 Merely Disjunctive Properties

It is a commonplace of the philosophy of causation that *merely disjunctive* properties, that is, properties formed by the disjunction of unrelated and heterogeneous properties, cannot be causally efficacious. For instance, one can explain a

fever by attributing the property of *having the mumps*, but not by attributing the property of *having the mumps or suffering from sunstroke*. The latter is not a natural property. Some disjunctions are not merely disjunctive: for example, the property of *being a marsupial or being a placental mammal* corresponds to the natural property of *being a mammal*.

The real difficulty lies in finding a principled way of distinguishing disjunctive predicates that represent merely disjunctive properties from those that represent natural ones. The most promising strategy is to make use of the causal laws (or type/type constraints) in which the property figures. A property $\phi \vee \psi$ is merely disjunctive just in case, for every property χ , the constraint $(\phi \vee \psi) | \sim \chi$ holds if and only if both $\phi | \sim \chi$ and $\psi | \sim \chi$ hold. However, this method of distinguishing merely disjunctive from natural properties fails in the present context, in which the constraints are held to be *deterministic*. Every disjunction would turn out, according to the deterministic model, to be merely disjunctive. This is one of the reasons for preferring the indeterministic and probabilistic models I develop in *Realism Regained*. For the purposes of the present paper, however, we shall simply have to take the characteristic of being a *natural type* (a type which is not *merely* disjunctive or *merely* general, intuitively speaking) as an undefined theoretical primitive.

8 Humean Supervenience and Singular Causation

Following Tooley, I will use ‘Humean supervenience’ to represent the thesis that the facts about token-causation supervene upon occurrent facts plus the facts about the actual causal laws. To deny Humean supervenience is to affirm the possibility of ‘singular causation’, causal connections whose existence is inexplicable in terms of causal laws and non-causal facts.

My account so far is neutral on the question of Humean supervenience. However, it does treat singular causal connection as a more basic notion than that of causal law. This does not preclude Humean supervenience, but it certainly makes this thesis an unnatural assumption to make without corroborating evidence.

In fact, the notion of ‘causal law’ does not play a central role in my account, in contrast to the Armstrong/Tooley tradition. I prefer to make use of modal and stochastic notions, rather than talking directly about “lawlike” generalizations.

If the hypothesis of the completeness of causal explanation is true, then every instance of token-level causation falls under some necessary generalization at the level of types. This implication of explanatory completeness is important enough to warrant separate attention. I shall refer to it as *Hume’s Hypothesis*, since its truth is entailed by the extensional adequacy of Hume’s definition of causation in terms of relations between types.

Hypothesis 1 (Hume’s Hypothesis) *If $(s \triangleright s')$ and $(s' \equiv \psi)$, then there*

exists a type ϕ such that $(s|\equiv (\phi|\sim \psi))$ and $(s|\equiv \phi)$.

Hume's Hypothesis can be generalized by use of the generalized (token-set) causation.

Hypothesis 2 (Generalized Hume's Hypothesis) *If $(s \triangleright B)$ and $\forall s' \in B(s'|\equiv \psi)$, then there exists a type ϕ such that $(s|\equiv (\phi|\sim \psi))$ and $(s|\equiv \phi)$.*

These hypotheses do not entail Humean supervenience, however, since even if they held, it could still be the case that which token is causally connected to which is *not* determined by the combination of non-causal facts about the tokens plus the type-level necessities. It may be that law-like generalizations always presuppose some irreducible facts about token-level causal connections. This is especially plausible if space and time are themselves constructible from such token-level causal connections. Typically, causal generalizations will make reference to the spatiotemporal relations between the cause and effect.

Both Armstrong and Tooley are quite exercised over the issue of whether causal laws are contingent or necessary (they both insist that these laws are contingent). Are necessities of causal connection themselves necessary or contingent? This is a familiar question in modal logic. It amounts to asking whether metaphysical necessity is at least *S4*, that is, is the relevant accessibility relation transitive? Armstrong and Tooley are, in effect, asserting that necessity is not *S4*, that some necessities are themselves non-necessary. I am inclined to believe that most causal necessities at least are contingent, but, unlike Armstrong and Tooley, I do not see any interesting metaphysical issues turning on this question.

Armstrong and Tooley seem to have a tendency to confuse the necessary/contingent contrast with the analytic/synthetic distinction. They seem to suppose that, if some causal laws were necessary, they would have to be analytic as well. Since no causal law is analytic, they infer that all causal laws are contingent. However, I cannot see how we can exclude the possibility that at least some causal laws are necessary but synthetic.

9 Empiricism and Modality

Van Fraassen has argued that the sort of naive reliance on modality that characterizes my approach violates certain empiricist strictures. In particular, van Fraassen argues that a modal realist like myself, who denies that modal facts supervene on the non-modal facts, cannot solve the "inference problem" [31]. This inference problem concerns the rationality of accepting axiom *T* of modal logic: if necessarily ϕ , then ϕ . Since I decline any attempt to define necessity, I cannot argue that *T* is an analytic truth, derivable by deductive logic from a set of stipulative definitions. How then can I claim that acceptance of *T* is rationally obligatory? If I deny that it is rationally obligatory, I have no basis for claiming that causal explanation is sound, or that causal necessities constrain the actual sequence of events in the world.

My response to van Fraassen is simply to insist that the acceptance of T is required by the proper functioning of the human mind, which I do not take to be exhausted by conformity to the demands of deductive logic. Axiom T is in fact always true, and necessarily so. Hence, reliance on T is highly reliable, as reliable as reliance on any axiom of standard first-order logic. The “inference problem” is a problem only for one who, like van Fraassen, is wedded to the Humean doctrine that the only standard of rational belief is closure under standard deductive logic.¹

10 Causal Relevance

A key notion in my definitions of teleo-function and of modal knowledge is that of *causal relevance*. There are two ways to define the causal relevance of the type of a token to a type of a second token. The first way makes use of the INUS connective, \rightsquigarrow .

Definition 10.1 (Causal Relevance, I) $(s : \phi) \rightsquigarrow (s' : \psi)$ if and only if (i) $s \rightsquigarrow s'$, (ii) $(s \equiv \phi)$ and $(s' \equiv \psi)$, and (iii) for all s'' , if $s \rightsquigarrow s''$ and $s'' \sqsubseteq s'$, then $s' = s''$.

In other words, $(s : \phi)$ is causally relevant to $(s' : \psi)$ just in case: $s \models \phi$, $s' \models \psi$, and s' is a minimal token verifying the relation $s \rightsquigarrow s'$. Thus, mereological minimality comes into the definition of causal relevance twice: first in the definition of the INUS condition (s is an INUS cause of s' just in case s is part of a minimal total cause of s'), and second, in the definition of causal relevance itself.

A second approach to the definition of causal relevance would be to define a relation of *subtype*. Type ϕ is a subtype of ψ just in case every possible token that verifies ϕ also verifies ψ . The intension of the subtype of a type is a subset of the type's own intension. Two types are identical if each is a subtype of the other, i.e., if their intensions coincide.

Using subtypes, we can define a *minimal explanation*:

Definition 10.2 (Minimal Explanation) $((s_1 : \phi) \triangleright_{min} (s_2 : \psi))$ if and only if ϕ is natural, and for every natural type χ such that $(s_1 : \chi)$ and ϕ is a subtype of χ , $((s_1 : \chi) \triangleright (s_2 : \psi))$ iff $\chi = \phi$.

Finally, causal relevance can be defined in terms of minimal explanation, exactly as INUS causation has been defined in terms of minimal token causation.

Definition 10.3 (Causal Relevance, II) $(s : \phi) \rightsquigarrow (s' : \psi)$ if and only if ϕ is natural, and there exists a type χ such that ϕ is a subtype of χ and $((s : \chi) \triangleright_{min} (s' : \psi))$.

It would be worthwhile to investigate under what conditions these two definitions of causal relevance coincide.

¹For a further discussion of this topic, as part of a general account of inductive knowledge, see Chapter 19 of *Realism Regained*. [18]

11 Modal and Nomic Facts as Causes

Modal facts can themselves act as causes. Suppose that s is a minimal cause of s' , that is, no proper part of s is a cause of s' . According to the definition of causation, s itself must support the modal fact $\Box(As \rightarrow As')$. Any part of s that does not support this modal fact must be a proper part, and so must not be a cause of s .

If we assume a principle of strict downward monotonicity, it follows that any type supported by a minimal cause of a token is causally relevant to any type supported by that token.

Hypothesis 3 (Strict Downward Monotonicity) *If $(s \triangleright s')$, and $s_1 \sqsubset s'$, then there exists an s_2 such that $s_2 \sqsubset s$ and $s_2 \triangleright s_1$.*

If $(s \triangleright B)$, and $C \sqsubset B$, then there exists an s' such that $s' \sqsubset s$ and $s \triangleright C$.

Strict downward monotonicity entails that if s is a minimal cause of s' , then s is not a minimal cause of any proper part of s' . If s is a minimal cause of s' , then it is certainly part of a minimal cause, and so s is an INUS cause of s' , $s \rightsquigarrow s'$. If strict downward monotonicity holds, then s is not an INUS cause of any proper part of s' . This means that $(s : \phi)$ is causally relevant to $(s' : \psi)$, for any ϕ and ψ such that $s \models \phi$ and $s' \models \psi$. In particular, in the case above, s 's being of the modal type $\Box(As \rightarrow As')$ is causally relevant to every type of s' .

Nomic facts can also be causally efficacious. In the case above, by the definition of \triangleright , s must support the causal constraint $s \sim s'$. If Hume's Hypothesis applies to this case, then there must be a type ϕ such that s supports both ϕ and the causal constraint $\phi \sim \psi$. By the definition of causal relevance, we have that the causal-constraint type $\phi \sim \psi$ supported by s is indeed causally relevant to the explanation of s' and its type ψ . The truth of the causal constraint at s is an indispensable part of the explanation of the actuality of an immediately posterior situation of type ψ .

To make this concrete, suppose that s is an event of the collision of a pair of billiard balls with specific velocities. The relevant physical type of s (representing the masses and velocities of the two balls, as well as their impenetrability and elasticity) is ϕ . The causal constraint $\phi \sim \psi$ is a special case of the laws of conservation of energy and momentum. This nomic fact is causally relevant to the subsequent velocities of the balls (represented by ψ). Since the type ψ is observable, our perceptual faculties belong to a causal chain including particular nomic facts. Such causal connections make possible reference to and knowledge of such laws of nature.

11.1 The Causal Relevance of the Law of Excluded Middle

If a token is of type $(\phi \vee \neg\phi)$, then it must be either of type ϕ or of type $\neg\phi$. In any given case, it is one type or the other that will be causally relevant. Merely disjunctive types are never causally relevant, since if $((\phi \vee \psi) \Box \rightarrow \chi)$ is a modal fact in a situation, then either $(\phi \Box \rightarrow \chi)$ or $(\psi \Box \rightarrow \chi)$ will also be facts.

Instances of the law of excluded middle are always heterogeneous disjunctions and, hence, never represent natural types. If we assume that one or the other of these is supported by some sub-token of the original, then the disjunctive law and the disjunctive property turn out to be causally irrelevant.

However, although $(\phi \vee \neg\phi)$ may never be causally relevant, the same cannot be said of the type $\Box(\phi \vee \neg\phi)$. Suppose that token s supports the following types:

- $\Box(\phi \vee \neg\phi)$
- $((\phi \& \chi)|\sim \psi)$
- $((\neg\phi \& \rho)|\sim \psi)$
- χ
- ρ

Let us assume that s does not support any other relevant types; in particular, let us assume that it does not support $((\neg\phi \& \chi)|\sim \psi)$, or $((\phi \& \rho)|\sim \psi)$. Token s does support the type $(\chi \& \rho)|\sim \psi$, since this follows from the first three types. However, let us assume that s supports $(\chi \& \rho)|\sim \psi$ only because it supports the first three types. That is, let us assume that any proper part s_1 of s that does not support all of the first three types above does not support $(\chi \& \rho)|\sim \psi$.

Given these types, it follows that s constrains the actuality of a succeeding token of type ψ . To be more precise, s must constrain the existence of a set B of types, each of which is immediately posterior to part of s and each of which supports the type ψ . If we assume, as seems reasonable, that s as a whole is causally prior to each member of B , it follows that s is a cause of B , $s \triangleright B$.

Under these assumptions, we can show that s is a minimal cause of B , if we also assume Hume's Hypothesis (the supervenience of token causation on type causation). Suppose that s' is a proper part of s , one that does not support one or more of the types listed above. Since s does not support any other relevant types, neither can s' , one of its proper parts.

Suppose, for example, that s' supports only the following four relevant types. It is easy to check, in a four-valued model, that these types are not sufficient to guarantee the actuality of a succeeding token of type ψ :

- $((\phi \& \chi)|\sim \psi)$
- $((\neg\phi \& \rho)|\sim \psi)$
- χ
- ρ

By our earlier assumption, s' does not support the type $(\chi \& \rho)|\sim \psi$. Let s'' be a situation accessible to s' that supports both χ and ρ , but is not succeeded by any token supporting ψ . Given the support by s' of the first three types

above, this entails that s'' falsifies both ϕ and $\neg\phi$ are falsified (i.e., both ϕ and $\neg\phi$ are supported by s''). This is possible, since s' does not support the modalized law of excluded middle. The existence of s'' demonstrates that s' cannot be a cause of B , since every member of B supports ψ . Consequently, s is a minimal cause of B .

As before, strict downward monotonicity entails that every type supported by s is causally relevant to every type supported by B , in particular, to type ψ .

Although I have made use, in this argument, of Hume's Hypothesis and the hypothesis of strict downward monotonicity, it is not essential to assume that these hypotheses hold universally. All that I need is that they hold in some cases of the appropriate kind.

For a concrete illustration, suppose that s represents a situation in which a rabbit is pursued by a pair of predators, χ representing the presence of predator P_1 and ρ representing the presence of predator P_2 . Let us suppose that predator P_1 does not yet perceive the rabbit, but will immediately perceive and devour the rabbit if the rabbit makes any sudden movement (ϕ). In contrast, predator P_2 has the rabbit within its perceptual field and will devour it *unless* the rabbit makes a sudden movement, in which case P_2 will lose track of the rabbit's location. The rabbit notices predator P_2 and, consequently, makes a sudden movement (ϕ), resulting in its demise ψ , in this case, due to the actions of predator P_1 .

My argument is that in this case, the situation s , which records the necessity of the disjunction $\phi \vee \neg\phi$, plus the two causal constraints, plus the facts χ and ρ , is in every sense a cause of the rabbit's demise, and the inclusion in s of the modalized logical truth is causally relevant to the result.

This result can be generalized to any validity of classical first-order logic, by simply substituting the validity for the Law of Excluded Middle, and adding causal constraints that interact appropriately with the logical validity.

12 The Causal Theory of Logical and Mathematical Knowledge

Gettier examples in epistemology indicate that a causal element is needed to distinguish knowledge from true opinion. The distinction between knowledge and true opinion pertains to the domain of logic and mathematics, as well as to domains of contingent and temporal truths. Hence, we need to be able to appeal to the existence of causal connections of the appropriate kind between mathematical facts and our beliefs about them. Paul Benacerraf has pressed this point as a basis for a critique of realist conception of mathematical truth: if mathematical facts are causally inert, we cannot know them.[2]

Additionally, a causal theory is needed to provide an account of how we are able to refer to particular mathematical objects. There are infinitely many mathematical structures that provide models of our theories of arithmetic: how, apart from a causal connection, are we to explain the fact that we refer to exactly

one of these structures in arithmetic? This point, like the last one, has roots in the work of Benacerraf, in particular, his “What Numbers Could Not Be”[3].

As soon as we try to do so, however, we face a dilemma. If we try to identify logical and mathematical facts with contingent, spatio-temporal facts, we distort the nature of mathematics and lose that which distinguishes it from other sciences, as is illustrated by John Stuart Mill’s version of mathematical empiricism. Alternatively, if we locate mathematical fact in a timeless, necessary Platonic heaven, we face the daunting task of finding a ladder to make possible commerce between the Platonic heaven and cognitive states on earth. Merely talking about “a priori knowledge”, or vague allusions to a special capacity of sight or touch – seeing that $2 + 2 = 4$, or grasping a mathematical truth – fail to give us the kind of substantive theory capable of sustaining the knowledge/opinion distinction.

I will attack one horn of the dilemma, the horn which has rarely been challenged². I will argue that numbers and other mathematical objects are real and are causally effective. Information about mathematical objects is conveyed to us causally, not by some mysterious faculty of “mathematical intuition”, but through our interactions (both sensory and active) with ordinary, everyday situations. My view might be described as a kind of naturalistic Platonism, as opposed to the mystical Platonism of those who postulate a mysterious, non-natural channel to the mathematical as the unique possession of the human mind.

12.1 Logico-modal Facts as Causes

I suggest that it is modal facts that provide Jacob’s ladder between temporal events and Platonic truths. In this paper, I have developed an account of causation according to which modal facts can act as causes of contingent, temporal events. Logical and mathematical facts have causal efficacy in their modalized forms. For example, consider the Law of Excluded Middle. If a token is of type $\phi \vee \neg\phi$, then it must be either of type ϕ or of type $\neg\phi$, since the relation between tokens and types is governed by the extension of the strong Kleene truth tables to four-valued logic (the Dunn-Belnap tables). The purely disjunctive type never figures as such in any causal chain. However, a situation token can be of the type $\Box(\phi \vee \neg\phi)$, without being of type ϕ or $\neg\phi$. As we have seen, this modalized disjunction can figure significantly in causal chains.

The existence of this vertical causal connection does not exclude the simultaneous existence of an ordinary, horizontal connection. To return to the example above, there will be in the world a situation s' that is either of type ϕ or of type $\neg\phi$. If this situation s' also supports the relevant causal law, then it will constitute a total cause of the existence of a situation of type ψ . This will not, however, be a case of overdetermination, since the two causal connections occur at different levels. The vertical causal connection, involving the modal property $\Box(\phi \vee \neg\phi)$ presupposes the existence of some horizontal connection involving

²The sole exception I know of is Penelope Maddy’s work.

either ϕ or $\neg\phi$.

I am concerned in this paper with our knowledge of logical and mathematical truths, as objects of belief, with our knowing *that* such-and-such is so. There is the additional fact that we know how to reason deductively and mathematically. Our knowing-how to reason is easier to explain than our knowing-that certain mathematical facts obtain. We learn to conform to certain patterns of inference through trial-and-error experience, teaching us which patterns are reliably truth-preserving.

My account of logical knowledge depends on two things: postulating the existence of logically complex situation-types (negative types, disjunctive types, etc.), and postulating that the support relation between tokens and complex types is governed by a four-valued interpretation scheme, namely, the Dunn tables (and their extensions to the first-order case).

In some actual situations, the facts are partial: if ϕ represents the presence of hydrogen, then neither ϕ nor $\neg\phi$, its negation, may be supported in certain parts of the world (parts representing features other than chemical ones, for instance). There are no actual, nor even any possible, situation-tokens supporting both ϕ and $\neg\phi$, however, this impossibility is itself a fact that may be supported in some situations and not in others.

In order to represent such modal partiality, it is convenient to use impossible, over-determined situation tokens in our models. I am a realist about modality, but (unlike David Lewis) not about possible-but-not-actual situations. What is really possible is the realization of a certain type: it is convenient to model this fact through the use of fictional objects such as possible-but-not-actual situation tokens. Similarly, in order to model modal partiality, it is convenient to make use of the fiction of impossible situation-tokens.

One major advantage to the Dunn tables (as well as to their three-valued counterpart, the strong Kleene tables), for my purposes, is that no formula is true in every interpretation, or, in my case, no type is supported by every token in every model. There are non-trivial logical consequences in partial logic; for example, $\phi \& \psi$ entails $\psi \& \phi$. However, there are no logical validities in this logic, no conclusions that can be validly drawn from an empty set of premises.

This means that every classical validity (every type that is supported by every totally defined token) corresponds to a piece of modal information that may or may not be supported by a given token. If we slap a \Box in front of a classically valid type ϕ , we produce a piece of logical information that can enter into causal explanations of concrete events, including our own perceptions and beliefs.

It is important, for my purposes, to distinguish between logical knowing-how (knowing when it is proper to draw a particular inference) and logical knowing-that (knowing the necessity of a given classical validity). Knowing-how is to be defined in terms of a reliable disposition to draw only the correct (strong-Kleene valid) inferences, where this disposition has logical reliability as its proper function (in the sense of natural teleology). Knowing-that involves knowledge of particular modal facts, which entails the existence of an appropriate causal link between the particular fact known and the knowing of it.

12.2 Is Logic Factual?

I am claiming that the subject matter of logic is a domain of fact, specifically, of modal fact. There is a long tradition in philosophy, beginning at least with Hume, that divides truths into two categories: matters of fact and matters of the relation of ideas. Logic is the paradigmatic example of the second category.

Hume's distinction depends on the assumption that there can be no necessity in the world other than that which is projected on the world by some sort of psychological necessity. This in turn was based on Hume's sensationalist theory of concepts: since we have no sensation of necessity between external causes and effects, we can have no real concept of such a thing.

It is hard to see how Hume's distinction can be defended, since if there really is no necessity in the world, then there is no psychological necessity either, and hence no necessary relations of ideas. Conversely, once we acknowledge that some mental representations are possible and others are impossible, what principled reason do we have for extending this distinction to extra-mental event-types?

Another objection to a factual theory of logical truth proceeds in this way: whatever is factual is contingent, logical truths are not contingent, and therefore logic is not factual. I deny the first step: many modal facts (perhaps all of them, if the S5 axioms are sound) are necessary.

12.3 Logic as the Precondition of Thought

There is another distinction between logical and factual truths that could be contended for: that logical falsehoods are unintelligible, while merely factual falsehoods are intelligible. I agree that logical impossibilities are unintelligible, but I do not accept the further inference that this makes logic non-factual. There are certain facts, namely the necessary ones, that it is unintelligible to deny. I would not limit this to logical necessities: it is unintelligible to deny any necessity, whether this is physical or causal or some other sort. It is unintelligible to deny that water is water, and it is also unintelligible to deny that water is H_2O . There is a difference between the two: I learn that the one would-be conception is unintelligible by learning logic, and I learn that the other is unintelligible by learning chemistry. A mental representation can represent a real possibility (and so represent intelligibly) only if there is, in the realm of modal reality, a possible situation for the representation to be about. Which representations represent intelligibly is itself a factual matter, a matter to do with the modal structure of the world.

When we say that a logical falsehood is "unintelligible" or "incoherent", there are three things we might mean:

1. It literally cannot be thought.
2. Believing it would make one vulnerable to Dutch book strategies (in which it is possible to lose but impossible to win).

3. It is logically false.

The third sense of “incoherent” of course trivially distinguishes logical impossibilities from other impossibilities. I am not denying that the class of logical necessities forms a natural and interesting class; I am merely denying that the account of logical truth is radically different from the rest of semantics.

I would deny that logical falsehoods are incoherent in the first sense above. People do in fact believe logical falsehoods, and this is an important and causally relevant fact about them. I agree that believing logical falsehoods makes one vulnerable to Dutch books, but so does believing any impossibility. The incoherency comes from believing the impossible, not the illogical.

In any case, even if logic is in some special sense a precondition of all thought, this fact is irrelevant to the project of explaining the possibility of thinking about and knowing logical truths. If logic is a precondition of all thought, this may give me a reason to think logically, but it does not (by itself) explain how it is that I know logical truths, or what it is that I am talking about when I am doing logic.

12.4 The Apriority and Unrevisability of Logic

Although I am defending a *causal* theory of logical and mathematical knowledge, it does not follow that I am committed to an *empiricist* account, à la John Stuart Mill. It is quite possible, and I think probable, that elementary logic and mathematics are knowable a priori, and, moreover, that they are in fact unrevisable, hard-wired into our minds. My point is that the content and the epistemic status of such a priori convictions stands in need of some kind of causal explanation.

The faculty of imagination plays a critical role in the acquisition of new logical and mathematical knowledge. I can discover that the sum of five and seven is twelve, even though I have never encountered twelve identifiable things in one setting. I can imagine two disjoint sets, one of five and the other of seven, and discover that their union must consist of twelve individuals. No manipulation of physical objects is needed.

Nonetheless, we can ask: how is it that such features of our faculty of imagination are knowledge-conferring? It must have something to do with the origin of the human mind, whether Darwinistically or otherwise. The formation of our faculty of imagination must somehow have been influenced by the relevant logical and mathematical facts, perhaps as these facts were causally efficacious in various episodes in our evolutionary history.

12.5 Logical and Physical Necessity

Heretofore I have emphasized the similarities between logical and physical necessity. Both are knowable via their causal influence on sequences of concrete events. Nonetheless, there are clearly different forms of modality: logical and

merely physical, to take two examples. It is physically impossible, but logically possible, that I should travel faster than the speed of light. Can I give an account of the difference between the two?

It is important in this context to be very clear about what sort of thing is it to which we are attributing possibility or impossibility. For example, is it a situation-token, a situation-type, or a proposition (the combination of a token or tokens with a type)? As an actualist, I believe that the only real tokens are actual ones. So, I view merely possible tokens as some sort of logical construction, built up from actual tokens and various situation-types. Such a construction represents a real possibility just in case the types involved have the modal property of being possibly instantiated (or possibly instantiated by or in a certain relation to certain actual tokens). Similarly, a proposition is possibly true just in case its type is possibly instantiated by its token. Thus, modality is primarily a category of property of situation-types.

A situation-type represents a logical possibility just in case some type logically isomorphic to it is possibly instantiated. (By *logically isomorphic*, I mean that one can be transformed into the other through the substitution of non-logical elements.) Dually, a situation-type represents a logical impossibility just in case no type logically isomorphic to it is possibly instantiated.

Analogously, a type constitutes a physical possibility just in case some type physically isomorphic to it is possibly instantiated. (Physical isomorphism means that one can transform one into the other by substitution of non-physical-type elements.) We normally include logical structure in our definition of physical structure, resulting in the inclusion of all physical possibilities within the class of logical possibilities. However, we need not do so: we could countenance certain types as physically possible but logically impossible. For example, it is physically possible for an electron to have spin $+1/2$, and it is physically possible for it not to have spin $+1/2$. We could count the type according to which the electron both has and does not have spin $+1/2$ as physically possible, but logically impossible.

My point is that possibility and necessity tout court are the basic realities. Logical modality and physical modality are two kinds of structure we find within the modal reality of the world. They are distinct, but not fundamentally different in kind.

It may be that there is a further difference between logical and physical necessity. It may be that physical laws are only contingently necessary, while the truths of logic are necessarily necessary. This could happen if we find that the laws of physics are themselves the resultant of some more fundamental fact (such as the will of God), while the laws of logic (or some of them) are absolutely underived.

A standard distinction between logical and non-logical necessity relies on Tarski's reduction of logical necessity to 'truth in every model'. The inadequacy of such a model-theoretic approach to logical necessity can be seen by considering propositional logic and truth tables. Suppose we try to identify logical truth in propositional logic with *true in every interpretation*, with the interpretations of the logical connectives simply stipulated by displaying the

corresponding truth functions. This theory of logical truth can work only by asserting (if only implicitly) that the rows of the truth tables are necessarily exclusive and exhaustive of all possibilities. This is something that cannot be simply stipulated to be the case.

For example, consider just negation. If by ‘false’, we mean ‘not true’, then the fact that the two rows of the standard truth table for negation are exclusive and exhaustive is itself a prior logical necessity, and not simply the product of our stipulating a meaning for ‘not’. Alternatively, if ‘false’ does not simply mean ‘not true’, then the standard truth table presupposes a substantive thesis of bivalence. In this case also, the mutually exclusive and exhaustive nature of the rows is not merely a product of convention. The semantic fact of bivalence is now something with which we must have some kind of epistemic contact, and this fact of bivalence is itself *modal* in character: we need to know, not only that every sentence in a certain class is in fact either true or false and not both, but that this holds of necessity. Once again, we encounter a modal fact to which we must have epistemic access.

12.6 From Logic to Arithmetic

When compared to our knowledge of logic, our knowledge of arithmetic poses a new challenge. Arithmetic involves the existence and properties of things, the numbers, that seem to exist in a realm causally isolated from our own. However, this appearance may be deceiving. A number is simply a natural kind of quantifier complex.³ Numbers and their properties are thereby contained in modalized logical facts. Whenever a situation supports a modalized logical fact involving quantifiers and identity, that situation also supports an arithmetical fact involving one or more numbers. For example, the logical type:

$$\Box[\exists x(\neg A(x) \wedge B(x)) \wedge \exists y(A(y) \wedge B(y)) \rightarrow \exists z\exists w(B(z) \wedge B(w) \wedge z \neq w)]$$

corresponds to the arithmetical type $1 + 1 \geq 2$. The number n is simply a kind of quantifier complex occurring in modalized logical facts, a complex consisting of n quantifiers whose variables are declared to be pairwise distinct. For instance, the following type is of type 3:

$$\exists x\exists y\exists z[x \neq y \ \& \ y \neq z \ \& \ x \neq z \ \& \ \phi]$$

The existence of logically complex types of this kind is not the result of any human doing. Our capacity to speak a recursive language and to think thoughts of arbitrary logical complexity all depends on the prior existence, in reality, of corresponding logical complexity. The commitment to an infinity of numbers and the commitment to the recursive nature of language are essentially the same things, as Gödel and Poincaré long ago realized. If we believe in the existence of a recursively defined language containing quantifiers and identity,

³Alternatively, it may be that each number is the causal ground of the existence of a family of equinumerous quantifier complexes.

we have already accepted the existence of the number series, since each number is simply a kind of quantifier complex producible in such a language.

In the ancient world, the Pythagoreans and the Eleatic philosophers argued over which was more fundamental: numbers or logic. As T. K. Seung has argued,[28, pages 194-5] the later Plato reached the conclusion (expressed in his *Parmenides*) that the two are inseparable and interdependent. As soon as we admit into our logic formulas of arbitrary complexity, i.e., as soon as we recognize that we are working with a syntax and semantics that can be defined only *recursively*, we are already committed to the real existence of the natural numbers.

Thus, numbers do have causal influence on the world: they do so by figuring in modalized logical facts that constrain what can happen. To posit that every number has a successor is to hypothesize that there exist real modal constraints of this kind of arbitrary logical complexity.

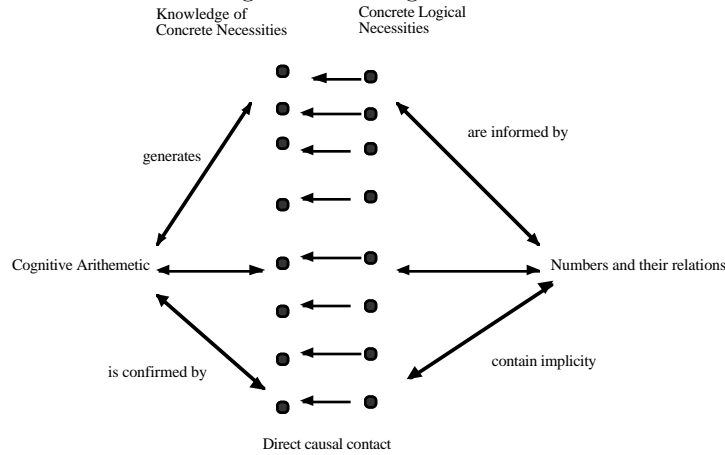
Thus, contrary to Hartry Field, arithmetic is not a conservative extension of physical theory. Rather, the axioms of arithmetic are an especially bold conjecture, a set of infinitary generalizations based on our knowledge of their instances. These arithmetical conjectures are confirmed every time we encounter novel situations of great complexity and are able to navigate through these situations successfully, with arithmetic as one of our guides.

12.7 Kripke and Wittgenstein on Rule-Following

Kripke finds in Wittgenstein's *Philosophical Investigations* a novel puzzle: how is it that a finite number of acts can fix the content of the rule being followed in a given practice.[19][32] In the case of arithmetic, the set of arithmetical calculations that ever have or ever will be performed is finite. There are infinitely many different extensions of these data points to the entire three-place Cartesian product of numerals. For example, the "quus" function differs from addition only on pairs of numbers so large that no one will ever use them. What makes it the case that we are following the rule of addition instead of its counterpart quaddition?

Kripke's puzzle seems to put the order of explanation the wrong way around. It is because we mean addition by "plus" that we are (or should be) following the addition rule, not vice versa. The fact that the linguistic and cognitive operations in question represent addition is determined by systematic causal connections between them and facts in the realm of logical necessity. Our arithmetical calculations are (teleologically speaking) *supposed to* connect in a particular way with the set of first-order logical necessities. Natural numbers can be systematically translated into strings of quantifiers, qualified with suitable identity or non-identity statements, as in Frege's logicist programme. Arithmetical calculation is supposed to facilitate efficient computation of logical necessities via these translations. If these cognitive operations represented quaddition instead, these systematic causal connections would be quite different (and a good deal more complicated).

I cannot think of any way of making sense of direct causal connections between bare mathematical facts (situations including only certain numbers and some mathematical relations between them, such as ‘ $7 + 5 = 12$ ’ or ‘ $3 < 5$ ’) and temporally-located events and processes. Instead, the connection is more indirect and holistic. Particular logical facts impinge directly on concrete events and processes. Implicit in these logical facts are numbers (types of quantifier complexes) and their mathematical relations (such as succession and inclusion). Representations of numbers and their relations in the mind (which we might call “cognitive arithmetic”) is confirmed by its reliability and fruitfulness in generating information about first-order logical necessities (via the translation of numbers into quantifiers restricted by identity and distinctness conditions). Thus, there are two systems, real arithmetic and cognitive arithmetic, whose agreement is caused and sustained by a finite number of causal interactions between first-order logical facts, facts about concrete necessities and possibilities, and cognitive facts constituting our knowledge of these modal facts.



Arithmetical facts are knowable by virtue of a systematic translation between atomic facts about numbers (facts about the value of particular sums and products) and modal facts of first-order logical necessity. Each atomic fact about numbers can be mapped to a corresponding set of theorems of first-order logic (as in standard logicist treatments of arithmetic). However, isn't this systematic translation between arithmetic and logic itself a rule that can be quusified? The translation is an infinitary rule, but our actual mathematical practice concerns only finitely many instances of this translation scheme. Doesn't the Kripke/Wittgenstein problem arise at this point?

The answer to this deviant translation problem is to posit that the numbers really exist, and really participate in those modal situations to which the translation scheme links them. That is, the numbers 2, 3 and 5 are real constituents of the modal situation-token that supports the logical necessity of the translation of ‘ $2+3=5$ ’ into first-order logic. Moreover, since these modal situation-tokens enter into causal relations to ordinary events and processes, the individual numbers are also implicated in these causal relations.

However, we still must confront the fact that we humans have interacted in this way with only finitely many natural numbers. What determines the extension of the translation scheme to numbers with which we have had no such interaction? The answer would have to take something like the following form. The successor relation is a real, quasi-causal relation between numbers. The successor relation enters into the very being of all numbers other than 0. Thus, in interacting with numbers larger than 0, we are interacting with the successor relation itself. Once we reach the point of forming universal generalizations about numbers, such as the generalization that every number has a successor, our representational state is informed by a causal connection to this successor relation (as a constituent of the relevant modal facts). In other words, succession is a natural kind, like *gold* or *aardvark*.

This is not to deny that quosified successor-like relations exist. It is merely to deny that these quosified relations enter into causal connections with our epistemic states when we form generalizations about numbers and their successors. Moreover, there are important causal/explanatory asymmetries between succession and quosuccession. The fact that every number has a successor explains why every number has a quosuccessor, but not vice versa.

12.8 Set Theory and Other Branches of Mathematics

Not all of the branches of mathematics will succumb to the same treatment as does arithmetic. In the case of geometry and real analysis, for instance, a structuralist theory along the lines of Dedekind seems called for[8]: Euclidean geometry is about any structure that satisfies its axioms, and similarly for non-Euclidean geometries and real analysis. The sort of “elimination” of real numbers proposed by Hartry Field[11] fits into this structuralist pattern: physics postulates the existence of various systems of physical quantities (distance, duration, mass, field intensity) which validate the axioms of real analysis (under suitable interpretation). Thus, physics, and other empirical sciences, verify the existence of structures of certain kinds, and the various branches of structuralist mathematics investigate the logical consequences of being structures of the postulated kind.

Consequently, there is no such thing as *the* real numbers, or Euclidean space, as subsistent objects. In real analysis, we do not study the properties and relations of a collection of objects (the real numbers); instead, we study the properties of any of a class of structures, each of which validates the axioms of analysis. In contrast, arithmetic is primarily the study of the natural numbers, which are definite objects in their own right, although, secondarily, its results can be applied to any omega-sequence (that is, any sequence sharing the structure of the natural numbers). The difference between arithmetic and analysis lies in the tightness of the connection between arithmetic and logic, via the definability of finite cardinality in first-order predicate logic.

Set theory constitutes a more difficult case. I would lean toward classifying it with arithmetic (as having an absolute domain, *the sets*), rather than with the structuralist branches, such as analysis and geometry. Just as numbers can

be thought of the grounds of natural families of logical types (types involving quantifier complexes of the corresponding cardinality), so to can sets be thought of as grounds of natural families of disjunctive logical types. For example, suppose the set A is $\{a_0, a_1, \dots, a_\alpha\}$. Set A is the cause of the existence of a family of co-extensive types, of which the following is an instance:

$$\lambda x(x = a_1 \vee x = a_2 \vee \dots x = a_\alpha)$$

Let A^\vee be the family of types co-extensive with this one. We can define membership as follows:

$$\forall x(x \in A \Leftrightarrow (x \equiv A^\vee))$$

Sets, even infinite sets, are essentially things that can be given by exhaustive enumeration. For this reason, something like Zermelo- Frankel set theory must be correct, since an enumeration can include only things that exist prior to the enumeration, a restriction reflected in the iterative hierarchy of Zermelo-Frankel theory.

How is it that we are influenced causally by sets and their mathematical properties? Quantum mechanics may play a key role in elucidating this connection. As David Bohm and Basil Hiley have argued, quantum mechanics concerns the process by which physical wholes are formed and dissolved.[6] Thus, in quantum mechanics, it is sets of physical objects, and not just objects taken individually, that are causally efficacious.

The possibility of the formation of physical wholes presupposes a prior metaphysical composition of individuals into sets. When a set of physical systems meets certain conditions, it constitutes a quantum whole, whose properties are not decomposable into the properties of the parts. Our experience of this process of physical gives evidence of an underlying metaphysical process by which arbitrary collections of things constitute a metaphysical whole, i.e., a *set*. Even a nominalist like Field, for instance, confidently makes use of an axiom asserting the existence of arbitrary sums of spacetime regions.[12, page 214]

Presumably, it this feature of quantum mechanics that is ultimately responsible for the Gestalt phenomenon in perceptual psychology. The formation of a quantum whole, comprising a system of perceptual objects and the perceiver's sensory system, is a precondition of the perception of a holistic Gestalt. Gestalt perception is thus literally the perception of sets of objects, and not just of the objects individually.

Penelope Maddy has argued for this latter point: the perceptibility of sets of physical objects.[22] ⁴ However, this perceptibility is not sufficient to provide an account of our knowledge of set theory. We need to have some explanation for our knowledge of general facts about sets, of the kind represented by the axioms of set theory. In addition, our perception of sets of physical objects does

⁴Maddy argues that set theory is metaphysically prior to arithmetic, since she postulates that numbers are properties of sets.[22, page 87] However, the truths of arithmetic are already implicit in first-order logic with identity, even without any apparent ontological commitment to the existence of sets. (See Field's "Reply to Maddy" [12, pages 208-9].)

not seem to have any direct bearing on our knowledge of the higher ranks of set theory, or of the existence of unit sets or the null set.

I think the solution to these problems can be found by thinking of our knowledge of sets of higher ranks as analogous to our knowledge of the future. We know the future, not by virtue of any effect of the future upon our minds, but through the influence on our minds and on the future by certain common causes. Similarly, our experience of the metaphysical combination of physical object into sets, without limitation, provides us with access to a general tendency in things toward agglomeration. This same process of agglomeration is also responsible for the formation of arbitrary sets of higher ranks. Our knowledge of the axioms of powersets, separation, choice and replacement all reflect our knowledge of the universality of this agglomeration process. The iterative hierarchy of Zermelo-Frankel set theory reflects a genuinely causal process by which new sets, of higher and higher rank, are generated through the successive agglomeration of sets of lower rank. (This generation of successive ranks does not take place in time, but in an order orthogonal to our time line.) Thus, we know that sets of sets exist, even though we have no direct contact with such things, since the very same tendency of individuals to form sets is at work, both in the physical world and in the realm of sets of higher ranks.

This view of mathematical ontology and epistemology has no revisionist implications for the practice of mathematics, unlike many philosophical positions, such as intuitionism, finitism, or modal-structuralism. Mathematics is a healthy, progressive science, exploring the structure of modal reality in much the same way as physics explores the structure of physical forces and interactions. Mathematicians need no help from philosophers to do their job.

However, this version of set-theoretic realism does not by itself settle the question of bivalence. A proposition in the language of set theory is determined to be true if, it is true in the universe limited to a particular rank in the iterative hierarchy, and its truth is necessarily preserved by the process of agglomeration that leads to still higher ranks. A proposition is determined to be false when its negation becomes determinately false. There could, in principle, be propositions whose truth value never stabilizes in this way but instead fluctuates from truth to falsity and back again as the ranks accumulate.

13 Conclusion

An internalist model of causation, one permitting *vertical causation* from modal facts to physical facts, opens the way to a novel form of Platonism, a Platonism that makes no appeal to mysterious faculties of intuition or acquaintance.

In this paper, I have concentrated on the relevance of causal internalism for logical and mathematical knowledge. However, I believe that it is equally important in providing an adequate account of our knowledge of the laws of nature and of natural or physical necessity. Unless the modal facts encoded in our formulations of the natural law are causally efficacious, it must remain mysterious how our thought and knowledge manage to make real contact with

these laws, and some form of scientific anti-realism seems inevitable.

References

- [1] David Armstrong. *A Materialist Theory of Mind*. Cambridge University Press, London, 1968.
- [2] Paul Benacerraf. Mathematical truth. In Paul Benacerraf and Hilary Putnam, editors, *Philosophy of Mathematics*, pages 403–420. Cambridge University Press, Cambridge, United Kingdom, 1983.
- [3] Paul Benacerraf. What numbers could not be. In Paul Benacerraf and Hilary Putnam, editors, *Philosophy of Mathematics*, pages 272–94. Cambridge University Press, Cambridge, United Kingdom, 1983.
- [4] G. Birkhoff. *Lattice Theory*. American Mathematical Society, Providence, 1940.
- [5] S. Blamey. Partial logic. In D. M. Gabbay and F. Guenther, editors, *Handbook of Philosophical Logic, Volume III*, pages 1–70. D. Reidel, Dordrecht, Holland, 1986.
- [6] David Bohm and Basil J. Hiley. *The Undivided Universe: an ontological interpretation of quantum theory*. Routledge, New York, 1993.
- [7] B. A. Davey and H. A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, 1990.
- [8] Richard Dedekind. *Was Sind und Sollen die Zahlen?* Vieweg, Braunschweig, 1888.
- [9] Fred I. Dretske. *Knowledge and the Flow of Information*. The MIT Press, Cambridge, Mass., 1981.
- [10] J. M. Dunn. Intuitive semantics for first-degree entailments and ‘coupled trees’. *Philosophical Studies*, 29:149–168, 1976.
- [11] Hartry Field. *Science without Numbers*. Princeton University Press, Princeton, 1980.
- [12] Hartry Field. Reply to maddy. *Pacific Philosophical Quarterly*, 71:206–222, 1990.
- [13] Jerry Fodor. *A Theory of Content and Other Essays*. The MIT Press, Cambridge, Mass., 1990.
- [14] Alvin Goldman. What is justified belief? In George Pappas, editor, *Justification and Knowledge: New Studies in Epistemology*. D. Reidel, Dordrecht, 1979.

- [15] G. M. Hardegree. An approach to the logic of natural kinds. *Pacific Philosophical Quarterly*, 63:122–132, 1982.
- [16] Terence Horgan. Mental quausion. *Philosophical Perspectives*, 3:47–76, 1989.
- [17] Robert C. Koons. Teleology as higher-order causation: A situation-theoretic account. *Minds and Machines*, 8:559–585, 1998.
- [18] Robert C. Koons. *Realism Regained: An Exact Theory of Causation, Teleology and the Mind*. Oxford University Press, New York, 1999.
- [19] Saul Kripke. *Wittgenstein on Rules and Private Language*. Harvard University Press, Cambridge, Mass., 1982.
- [20] Ernest LePore and Barry Loewer. More on making mind matter. *Philosophical Topics*, 17:175–91, 1989.
- [21] David Lewis. *Counterfactuals*. Blackwell, Oxford, 1973.
- [22] Penelope Maddy. *Realism in Mathematics*. Clarendon Press, Oxford, 1990.
- [23] R. Muskens. *Meaning and Partiality*. Center for the Study of Language and Information, Stanford, California, 1995.
- [24] Alvin Plantinga. *Warrant and Proper Function*. Oxford University Press, Oxford, 1993.
- [25] John Pollock. *Contemporary Theories of Knowledge*. Rowman and Littlefield, Savage, Maryland, 1986.
- [26] Hilary Putnam. *Mind, Language and Reality*. Harvard University Press, Cambridge, Mass., 1975.
- [27] Wesley Salmon. *Scientific Explanation and the Causal Structure of the World*. Princeton University Press, Princeton, NJ, 1984.
- [28] T. K. Seung. *Plato Rediscovered: human value and social order*. Rowman and Littlefield, Lanham, MD, 1996.
- [29] Charles Taylor. *The Explanation of Behavior*. Routledge & Kegan Paul, 1964.
- [30] E. G. C. Thijsse. *Partial Logic and Knowledge Representation*. Eburon Publisher, Delft, the Netherlands, 1992.
- [31] Bas van Fraassen. *Laws and Symmetry*. Oxford University Press, Oxford, 1987.
- [32] Ludwig Wittgenstein. *Philosophical Investigations*. Macmillan, New York, 1953.
- [33] Larry Wright. *Teleological Explanations*. University of California, Berkeley, 1976.